CIS 419/519: Homework 4

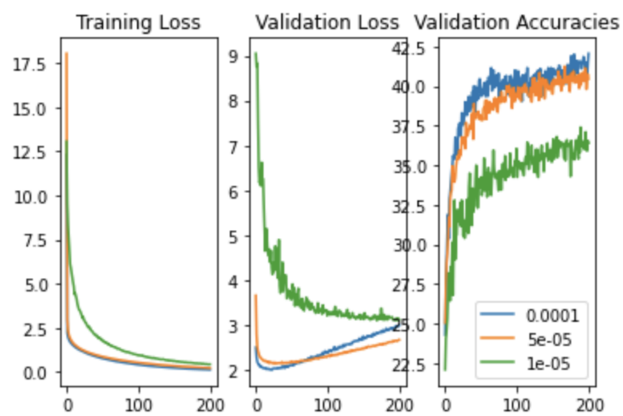
{Shall Chee Shih}

Although the solutions are my own, I consulted with the following people while working on this homework: {Names here}

Neural Networks

Feed Forward

Plots:

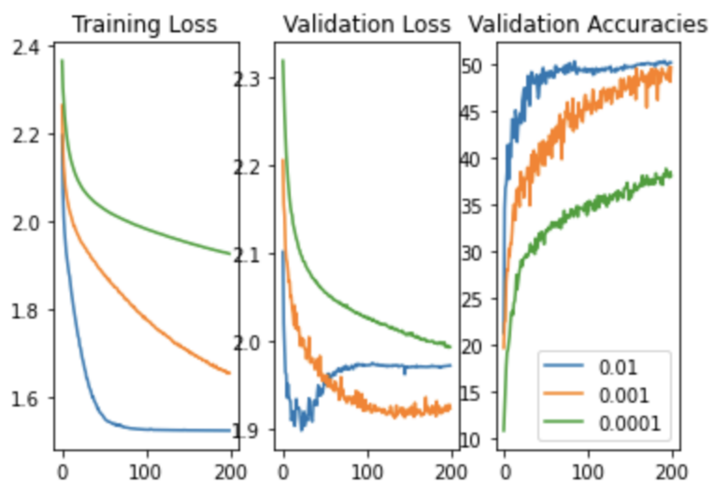


Learning Rate	Final Validation Accuracy
0.0001	42.05
0.00005	40.45
0.00001	36.4

Learning rate with best validation accuracy: 0.0001. The test accuracy of that model is 40.05%.

Convolutional

Plots:

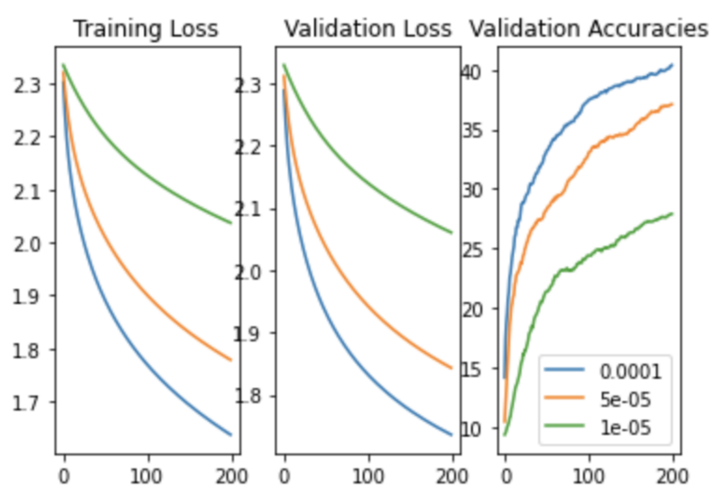


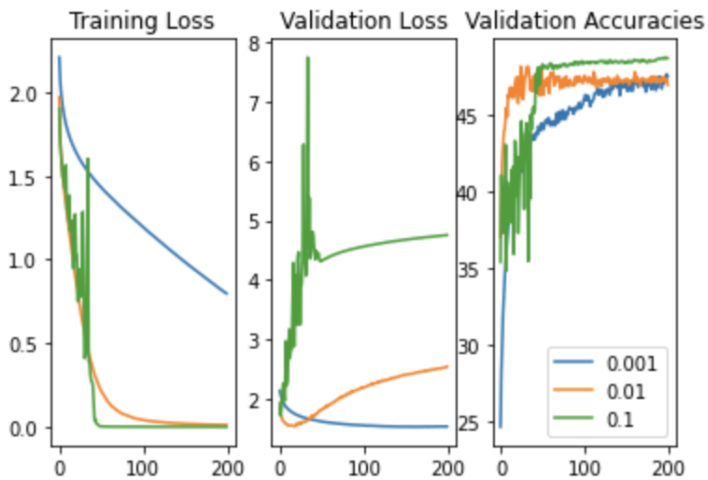
Learning Rate	Final Validation Accuracy
0.01	50.15
0.001	49.65
0.0001	38.05

Learning rate with best validation accuracy: 0.01. The test accuracy of that model is 49.2%.

Normalization : Feed Forward

Plots:



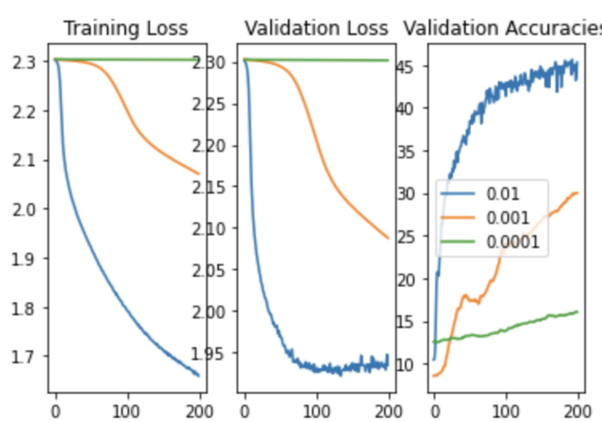


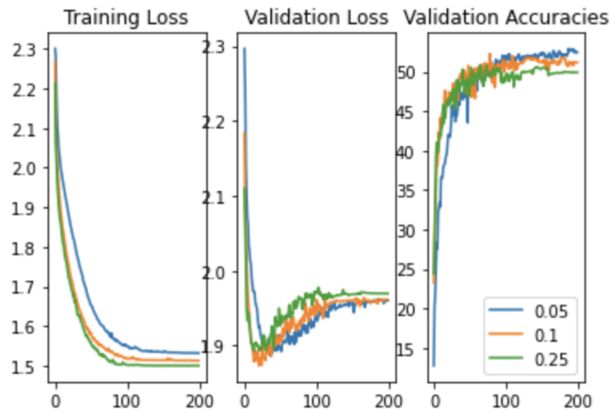
Learning Rate	Final Validation Accuracy
0.0001	40.4
0.00005	37.1
0.00001	27.9
0.001	47.5
0.01	46.95
0.1	48.7

Learning rate with best validation accuracy: 0.1. The test accuracy of that model is 45.05%.

Normalization : Convolutional

Plots:





Learning Rate	Final Validation Accuracy
0.01	45.3
0.001	30.0
0.0001	16.05
0.05	52.45
0.1	51.15
0.25	49.9

Learning rate with best validation accuracy: 0.05. The test accuracy of that model is 50.2%.

I used learning rates {0.0001, 0.00005, 0.00001, 0.001, 0.01, 0.1} for Feedforward model, and the performance improves. The test accuracy improves from 40.05% to 45.05%.

I used learning rates {0.01, 0.001, 0.0001, 0.05, 0.1, 0.25} for Convolutional model, and the performance improves. The test accuracy improves from 49.2% to 50.2%.

Document Classification

Representation	k	Test Accuracy
BBoW	0.1	0.86
CBoW	0.1	0.8608
TF-IDF	1.0	0.83

What impact does k have? That is, as the value of k goes to infinity, what will happen to $P(y \mid d)$?

$$P(v \mid y) = \frac{k + \sum_{d \in \mathcal{D}: y_d = y} f_d(v)}{k \cdot |\mathcal{V}| + \sum_{w \in \mathcal{V}} \sum_{d \in \mathcal{D}: y_d = y} f_d(w)}$$

k is a hyper parameter which controls the strength of the smoothing. As shown in the equation, when the value of k goes to infinity, $P(v \mid y)$ goes to $1/|V|$. Then,

$$\begin{aligned} P(y \mid d) = \frac{P(d \mid y) \cdot P(y)}{P(d)} &= \frac{\prod_{v \in d} \frac{1}{|V|} \cdot P(y)}{P(d \mid y=0) \cdot P(y=0) + P(d \mid y=1) \cdot P(y=1)} \\ &= \frac{\prod_{v \in d} \frac{1}{|V|} \cdot P(y)}{\prod_{v \in d} \frac{1}{|V|} \cdot P(y=0) + \prod_{v \in d} \frac{1}{|V|} \cdot P(y=1)} = P(y) \end{aligned} \tag{1}$$

Therefore, as k goes to infinity, $P(y \mid d)$ goes to $P(y)$, the numbers of the documents in the training data and vocabulary that have label y have less influence on $P(y \mid d)$.

Theory

Multivariate Exponential Naive Bayes

1.

$P(Y{=}A) = 3/7$	$P(Y{=}B) = 4/7$
$\lambda_{A;1} = 3/8$	$\lambda_{B;1} = 4/18 = 2/9$
$\lambda_{A;2} = 3/19$	$\lambda_{B;2} = 4/14 = 2/7$

2.

$$\frac{3/8e^{-3/8 \cdot 3} \cdot 3/19e^{-3/19 \cdot 5}}{4/18e^{-4/18 \cdot 3} \cdot 4/14e^{-4/14 \cdot 5}} \tag{2}$$

3. if

$$\frac{P(X_1 \mid Y=A) \cdot P(X_2 \mid Y=A) \cdot P(Y=A)}{P(X_1 \mid Y=B) \cdot P(X_2 \mid Y=B) \cdot P(Y=B)} > 0 \tag{3}$$

then $Y = A$, else, $Y = B$ is more possible.

4.

$$\frac{3/8e^{-3/8 \cdot 3} \cdot 3/19e^{-3/19 \cdot 5} \cdot 3/7}{4/18e^{-4/18 \cdot 3} \cdot 4/14e^{-4/14 \cdot 5} \cdot 4/7} = 0.838 < 1 \tag{4}$$

Therefore, Y = B is more possible.

Coin Toss

The most likely value of p is $\sqrt{2/5}$ and here's why. $P(H) = 1 - p + p(1-p) = 1 - p^2$, $P(T) = p^2$. According to maximum likelihood estimation, we need to maximize

$$P = 6 \cdot log(1 - p^2) + 4 \cdot logp^2 \tag{5}$$

Then, let

$$\frac{dP}{dp} = 6 \cdot \frac{-2p}{1-p^2} + 4\frac{2p}{p^2} = 0 \tag{6}$$

Therefore, $p = \sqrt{2/5}$.